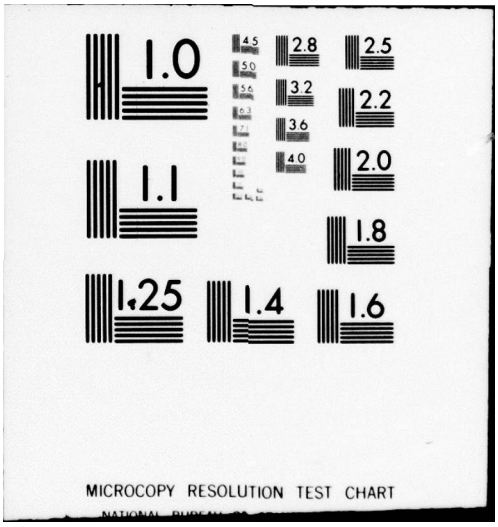
AD-A074 526 AEROSPACE GUIDANCE AND METROLOGY CENTER NEWARK AFS OH F/G 14/2
LASER POWER METER INTERCOMPARISON TEST EVALUATED BY TWO VARIABL--ETC(U)
SEP 78 W BLACKMON
UNCLASSIFIED AGMC-79-001-1 NL

| OF |
ADA
074526

END
DATE
FILMED
11 -79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART

LEVEL II

2

B.S.

LASER POWER METER INTERCOMPARISON TEST EVALUATED BY TWO VARIABLE LINEAR REGRESSION ANALYSES

AD A074526

Aerospace Guidance & Metrology Center
Newark Air Force Station
Newark, Ohio 43055

21 Sep 1978

DDC FILE COPY

DDC
RECEIVED
OCT 2 1979
D

Prepared by Willie Blackmon
WILLIE BLACKMON, Physicist

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

79 09 24 343

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (14) AGMC-79-001-1 / 2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER (9)
4. TITLE (and Subtitle) (6) Laser Power Meter Intercomparison Test Evaluated by two Variable Linear Regression Analysis.	5. TYPE OF REPORT & PERIOD COVERED Project Report Apr 1978 to Sep 1978
	6. PERFORMING ORG. REPORT NUMBER Project 1-3 (MAP)
7. AUTHOR(s) (10) Willie Blackmon	8. CONTRACT OR GRANT NUMBER(s) N/A
9. PERFORMING ORGANIZATION NAME AND ADDRESS Aerospace Guidance & Metrology Center Newark Air Force Station Newark, Ohio 43055	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS Aerospace Guidance & Metrology Center Directorate of Metrology - ML Applied Sciences Laboratory - MLLS (11)	12. REPORT DATE 21 September 1978
	13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same as above (12) 33	15. SECURITY CLASS. (of this report) Unclas
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

No restrictions on distribution

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

No restrictions on distribution

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Laser; power measurements

79 09 24 343

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Applied Science Laboratory under the Directorate of Metrology has a service function to provide laser power and energy measurement centrifications for all Air Force Agencies in need of such service. This report describes how the data collected in a laser power meter exchange program between the Directorate of Metrology and Brook AFB, Texas is analyzed in determining the errors in the ~~Brook's~~ intercomparison measurement process.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED 404 965

TABLE OF CONTENTS

	Page
INTRODUCTION	1
SUMMARY	2
DISCUSSION	3-17
CONCLUSIONS	18-19
REFERENCES	20
FIG. 1 - TIME PHASE DIAGRAM	21
FIG. 2 - CURVE PLOT	22
FIG. 3 - PROCESS EVALUATION DIAGRAM	23
A. MEASUREMENT DATA SHEET	24
B. PERCENTILE F - DISTRIBUTION	25
C. PERCENTILE t - DISTRIBUTION	26
D. REJECTION OF OUTLYING OBSERVATIONS	27-30

INTRODUCTION

The Applied Science Laboratory, AGMC, has a requirement to establish a CW Laser Power Meter Measurement Assurance Program between this laboratory and several Air Force Research and Development Laboratories.

This report describes and proposes a single and easy to apply criterion which allows one to objectively characterize and quantitatively evaluate a measurement assurance program of the type required to support the Air Force R&D Laboratories.

The criterion described is based upon a linear regression analysis of the output volts (X) from the standard power meter compared to the output volts (Y) of another power meter called the transfer standard. Each comparison between X and Y is treated as a coordinate pair (X,Y).

The analysis method used in this report follows that described in NBS Handbooks 91 and 300, Chapter 5-4, "Problems and Procedures for Functional Relationships". The chapter and title in both books are the same.

Accession For	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

DDC RECEIVED OCT 2 1979 D

SUMMARY

Using the voltage measurements from the standard and the transfer standard as X and Y respectively and treating X and Y as linearly related the following quantities were determined.

Mean $\bar{X}$	4.4756
Std Dev $\bar{X}$	.1073
Median X	4.45
Mean $\bar{Y}$	4.429
Std Dev $\bar{Y}$	.1237
Median Y	4.31
Intercept b_o	-0.64587
Slope b_1	1.13392
Variance in Y	.0005186
Variance in slope b_1	.0023691
Variance in intercept b_o	.047480
Correlation coefficient r	.98
Error in the slope b_1	.51%
Error in the intercept b_o	$\pm$2.29%
Whole line 95% confidence interval W_1	$\pm$.0382
Whole line 95% confidence error in W_1	$\pm$.93%
On line 95% confidence interval W_2	$\pm$.0301
On line 95% confidence error in W_2	$\pm$.73%
Single future 95% confidence interval W_3	$\pm$.0565
Single future 95% confidence error in W_3	$\pm$1.37%
Correlation coefficient r	.98

The resulting equation was determined to be:

$$Y = -0.64587+1.13392X$$

DISCUSSION

In this intercomparison process a laser power meter transfer standard was compared to a laboratory standard by alternately placing each into the laser beam. The voltage readings were taken digitally at the end of a 100 second exposure period. See Time Phase Diagram, Fig. 1.

Twenty voltage intercomparison measurements were made comparing the transfer standard voltage Y to the standard voltage X. In this analysis the X values were treated as independent variables and the Y values were treated as dependent variables. The measured values of X and Y were treated as 20 pairs of independent measurements, since the X values were measured on a different instrument from that of the Y values. Each comparison between X and Y is treated as a coordinate pair (X,Y).

A linear relationship is assumed between the two variables X and Y

$$Y = b_o + b_1 X \qquad (1)$$

The method of least square regression analysis is used for determining the linear relationship between the two variables X and Y (regression line of Y on X).

A rough plot of the values as coordinate pairs (X,Y) showed that they approximate a straight line.

For those who have access to computers or the hand calculators, the operations for determining the coefficients b_o and b_1 in equation (1) above can be made quite simple.

The approach given here is appropriate for the hand calculator. From the 20 pairs of values taken and shown on page 5 , the following was determined:

$$\Sigma X = 89.5100$$

$$\Sigma X^2 = 400.8209$$

$$\Sigma XY = 396.6880$$

$$\Sigma Y = 88.5800$$

$$n = 20.0000$$

$$\Sigma Y^2 = 392.6116$$

$$\bar{X} = 4.4755$$

$$\bar{Y} = 4.429$$

From the above values the necessary calculations were made and the results are shown on the worksheet, page 6. See appendix sheet A for the necessary formulae and instructions required to complete all calculations entered on the worksheet.

DATA SHEET

MEASURED VALUES

X VOLTS (STD)	Y VOLTS (TRANS STD)
4.54	4.54
4.57	4.51
4.56	4.50
4.50	4.46
4.48	4.45
4.46	4.42
4.45	4.39
4.66	4.64
4.68	4.66
4.67	4.64
4.43	4.41
4.42	4.41
4.45	4.38
4.43	4.37
4.42	4.36
4.36	4.28
4.36	4.28
4.31	4.21
4.37	4.34
4.39	4.33

BASIC WORKSHEET FOR ALL TYPES OF LINEAR RELATIONSHIPS

X denotes Voltage, standard Y denotes Voltage, transfer standard

ΣX = 89.5100 ΣY = 88.5800

$\bar{X}$ = 4.4755 $\bar{Y}$ = 4.4290

Number of points: n = 20

Step (1) ΣXY = 396.6880

(2) $(\Sigma X)(\Sigma Y)/n$ = 400.60201

(3) S_{xy} = .24821

(4) ΣX^2 = 400.8209

(5) $(\Sigma X)^2/n$ = 400.60201

(6) S_{xx} = .21890

(7) ΣY^2 = 392.6116

(8) $(\Sigma Y)^2/n$ = 392.32082

(9) S_{yy} = .29078

(10) $b_1 = \frac{S_{xy}}{S_{xx}}$ = 1.13392

(14) $\frac{(S_{xy})^2}{S_{xx}}$ = .28144

(11) $\bar{Y}$ = 4.429

(15) $(n - 2)s_y^2$ = .00934

(12) $b_1\bar{X}$ = 5.07487

(16) s_y^2 = .0005186

(13) $b_o = \bar{Y} - b_1\bar{X}$ = -.64587

s_y = .022773

Equation of the line:

$Y = b_o + b_1X$

$-.64587+1.13392X$

Estimated variance of the slope:

$s_{b_1}^2 = \frac{s_y^2}{S_{xx}}$ = .0023691

s_{b_1} = .048674

Estimated variance of intercept:

s_{b_o} = .21790

$s_{b_o}^2 = s_y^2 \frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}$ = .047480

Note: The following are algebraically identical:

$S_{xx} = \Sigma(X - \bar{X})^2$; $S_{yy} = \Sigma(Y - \bar{Y})^2$; $S_{xy} = \Sigma(X - \bar{X})(Y - \bar{Y})$.

Ordinarily, in hand computation, it is preferable to compute as shown in the steps above. Carry all decimal places obtainable--i.e., if data are recorded to two decimal places, carry four places in Steps (1) through (9) in order to avoid losing significant figures in subtraction.
"Copied from NBS Handbook 91, pg 5-10, U.S. Govt Printing Office, Washington, D.C."

The percent error % in the slope at the 95% confidence interval is determined. The student "t" value for 95% confidence and 18 degrees of freedom t(.95,18)= 2.101. See a copy of the table in appendix C "Percentiles of the "t" Distribution".

$$\% = \frac{t(.95,18)\ S_{b_1}}{\frac{n}{\bar{X}}} \times 100$$

$$= \frac{2.101(.048674)X100}{\frac{20}{4.4755}}$$

$$\% = \pm .51\%$$

Likewise the percent error % at the 95% confidence interval for the intercept

$$\% = \frac{t(.95,18)\ S_{b_o}}{\frac{n}{\bar{X}}} X\ 100$$

$$\% = \frac{2.101(.2178)}{\frac{20}{4.4755}} X\ 100$$

$$\% = \pm 2.29$$

Estimate W_l the 95% confidence interval for the whole line.

See Figure 2.

$$W_l = \sqrt{2F} \quad Sy \left[\frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}} \right] ½$$

F the percentile of distribution is taken from Table A-5 in the back of NBS Handbook 91. $F_{.95}$ (2,18) = 3.55. See Appendix B.

Determine W_l for several values X,(4.2<X<4.8).

X	$\pm W_l$	Y	$\pm$%
4.2	.0382	4.12	.93
4.25	.0322	4.17	
4.30	.0265	4.23	.63
4.35	.0212	4.29	
4.40	.0167	4.34	.38
4.45	.0140	4.40	
4.50	.0139	4.46	.31
4.55	.0167	4.51	
4.60	.0211	4.57	.46
4.65	.0264	4.63	
4.70	.0321	4.68	.69
4.75	.0381	4.74	
4.80	.0442	4.80	.92

The W_l interval $\pm$.0382 or $\pm$.93% is the widest interval falling within 4.2<X<4.8, the whole line range. Consequently this is the interval for which we are 95% certain that all values Y will fall for the whole line.

Estimate W_2 the 95% confidence interval for a single point on the line (i.e. the mean values of Y corresponding to chosen values of X). See Figure 2.

$$W_2 = t_{1-\alpha/2} \quad S_y \left[\frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}} \right]^{\frac{1}{2}}$$

$$= .05$$

$$1-\frac{\alpha}{2} = .975$$

The t (the percentile of distribution) is taken from Table A-4 in the NBS Handbook 91, $t_{.975}$ for 18 degrees of freedom is 2.101. See Appendix C.

Determine W_2 for values X, $(4.2<X<4.8)$.

X	$\pm W_2$		$\pm$%
4.20	.0301	4.12	.73
4.25	.0254		
4.30	.0209	4.23	.49
4.35	.0167		
4.40	.0132	4.34	.30
4.45	.0110		
4.50	.0110	4.46	.25
4.55	.0131		
4.60	.0166	4.57	.36
4.65	.0208		
4.70	.0253	4.68	.54
4.75	.0300		
4.80	.0349	4.8	.73

The widest interval for a point on the line is $\pm$.0349 or $\pm$.73%. Consequently we may state for a given value of X, that the mean value Y will fall within +.0349 or $\pm$.73% 95% of the time.

Estimate the 95% confidence interval W_3 for a single future value of Y for a single chosen value of X. See Figure 2.

$$W_3 = t_{1-\alpha/2}\, S_y \left[1 + \frac{1}{n} + \frac{(X-\bar{X})^2}{S_{xx}} \right]^{\frac{1}{2}}$$

The future values are taken to mean if the same measurement process is repeated in the future these future values of Y for chosen values of X are expected to fall within $\pm W_3$.

$$t_{.975} = 2.101$$

Determine several values of W_3 for chosen values of X, $(4.2<X<4.8)$.

X	$\pm W_3$	Y	$\pm$%
4.2	.0565	4.12	1.37
4.25	.0542	4.17	1.30
4.30	.0522	4.23	1.23
4.35	.0507	4.29	1.18
4.40	.0496	4.34	1.14
4.45	.0491	4.40	1.12
4.50	.0491	4.46	1.10
4.55	.0496	4.51	1.10
4.60	.0507	4.57	1.11
4.65	.0522	4.63	1.13
4.70	.0541	4.68	1.16
4.75	.0565	4.74	1.19
4.80	.0592	4.80	1.23

The W_3 interval $\pm$.0565 or $\pm$1.37% is the widest interval falling within $4.2<X<4.8$ the whole line range. Consequently future determinations of Y values using the least square fitting method for the same measurement process are expected to fall within this interval 95% of the time.

The degree of correlation between the variables X and Y is measured by the sample correlation coefficient r.

Determine the correlation coefficient r for the sample.

$$r = \frac{S_{xy}}{S_{xx} \quad S_{yy}}$$

$$= \frac{.24821}{(.21890) \quad (.29078)}$$

$$= .9838$$

$$r = .98$$

This method of analysis shows that the mean voltage of the transfer standard $\bar{Y}$ can be computed as a function of a given standard voltage X by the following equation:

$$\bar{Y} = -0.64587 + 1.13392X$$

This equation does not produce a single value of Y for a single value of X. It produces $\bar{Y}$ the mean of n Y values. One would not expect to make a single measurement getting the computed value $\bar{Y}$; one must make n measurements i.e. 5 measurements the mean of which would fall on the $\bar{Y}$ line. See Figure 2. To check this statement, go to $\bar{Y}$ values table pg. 12 and select 5 consecutive Y values i.e. 4.32,

CALCULATED VALUES OF $\overline{Y}$ IN TERMS OF X $\overline{Y} = -0.64587 + 1.13392X$				
	X	$\overline{Y}$	CONTINUED X	CONTINUED $\overline{Y}$
	4.30	4.23	4.54	4.50
	4.31	4.24	4.55	4.51
	4.32	4.25	4.56	4.52
	4.33	4.26	4.57	4.54
	4.34	4.28	4.58	4.55
	4.35	4.29	4.59	4.56
	4.36	4.30	4.60	4.57
	4.37	4.31	4.61	4.58
	4.38	4.32	4.62	4.59
	4.39	4.33	4.63	4.60
	4.40	4.34	4.64	4.62
	4.41	4.35	4.65	4.63
	4.42	4.37	4.66	4.64
	4.43	4.38	4.67	4.65
	4.44	4.39	4.68	4.66
	4.45	4.40	4.69	4.67
	4.46	4.41	4.70	4.68
	4.47	4.42		
	4.48	4.43		
	4.49	4.45		
	4.50	4.46		
	4.51	4.47		
	4.52	4.48		
	4.53	4.49		

4.33, 4.34, 4.35 and 4.37; the mean of this group is 4.34. The X value opposite 4.34 is 4.40. Consequently if one maintained the standard voltage X at 4.40 and made a single measurement it is equally likely than any one of the five Y value would occur.

To get a clear understanding of the criteria used in evaluation this intercomparison process it is important to correctly interpret the meaning of the values W_1, W_2 and W_3. There are errors inherent in any measurement process; consequently in this measurement, the process procedures, the operator's skill, the environmental conditions and the equipment used attribute to inherent errors which cause variation in the recorded voltages of the standard X and the transfer standard Y. These variations are expressed as $\pm W_1$, $\pm W_2$ and $\pm W_3$, and they fluctuate in the Y direction above and below the $\bar{Y}$ line. From the W_1 computations and Figure 2, it is noted that the value $\pm W_1$ is chosen, such that it envelops the worst deviation which occurs at X = 4.2 and X = 4.8. Consequently the process $\pm W_1$ variation is $\pm$.0382 volts above and below the $\bar{Y}$ line or $\pm$.93% above and below the $\bar{Y}$ line. It is noted from the computation also that W_1 decreases as values of X are held near the mean $\bar{X}$ = 4.476 which itself is limited by the inability of the laser to remain fixed at one power setting. W_2 defines the variation in the $\bar{Y}$ line itself. Consequently for any Y value located directly on the $\bar{Y}$ line itself, its worst variation $W_2 = \pm.0301$ or $\pm$.73% which occurs also at X = 4.2 and X = 4.8. Its minimum deviation also occurs around $\bar{X}$ = 4.476 similar to W_1.

W_3 defines the interval inside which future and subsequent measurement variations in Y are expected to fall. Consequently, for future and subsequent application of the same intercomparison process using the lease squares method, the predicted variations in W_3 are: $\pm$.0563 or $\pm$1.37% above and below the $\bar{Y}$ line.

It is noted that all values of W_1, W_2 and W_3 are computed at the 95% confidence interval. The use of the 95% or 99% confidence interval is arbitrary, however, since the 95% confidence interval is more frequently used, we have chosen to use it in our computations. By arbitrarily choosing the 95% confidence interval for our computations, this means that we are 95% confident that this measurement process will yield W values which will fall within the computed W_1, W_2 and W_3 intervals.

This also means that we have chosen to live with the risk that 5% of the W values are expected to fall outside the computed intervals.

Twenty comparison measurements were made during this test mainly because this was the first time this process was being evaluated. These data were collected over a period of two days. Now that the initial evaluation has been completed, future evaluations will be done by computing the W intervals based upon 10 data pairs X, Y.

Invariably after the data has been recorded, one is faced with the problem of what to do with one or two suspected values which will tend to shift the average away from that which is desired by the experimenter.

Now, in order to have a meaningful measurement program with credibility, one must adopt and accept some degree of objectivity and uniformity when it comes to the rejection of suspected data values.

The method suggested here is a simple one, so chosen to encourage its use, rather than a more complex method which would discourage its use.

The method applied here in testing whether to accept or reject a suspected value is described in NBS Handbook 300 Volume 1, p. 349-520 titled "Rejection of Outlying Observations". See instructions from this article Appendix D. p. 27.

First list the X and Y values in the order of size.

X	Y
4.31	4.21
4.36	4.28
4.36	4.28
4.37	4.33
4.39	4.34
4.42	4.36
4.42	4.37
4.43	4.38
4.43	4.39
4.45	4.41
median	
4.45	4.41
4.46	4.42

X	Y
4.48	4.45
4.50	4.46
4.54	4.50
4.56	4.51
4.57	4.54
4.66	4.64
4.67	4.64
4.68	4.66

Since we have 20 values the appropriate ratio is:

$$r_{22} = \frac{X_3 - X_1}{X_{n-2} - X_1} \qquad n = 14 \text{ to } 30$$

Referring to the above list of X values, $X_1 = 4.31$, $X_2 = 4.36$ and $X_{18} = 4.66$.

$$r_{22} = \frac{4.36 - 4.31}{4.66 - 4.31}$$

$$r_{22} = .14$$

Referring to Table I Appendix D, Testing for Extreme Observations, under the column marked = 5 which has the 95% confidence interval critical values, it is noted that at n = 20 the critical value is .45. Since the above computed value $r_{22} = .14$ is less than the critical value .45 the suspected value of 4.31 is not a mistake and may not be excluded. In like manner the decision is the same when the same procedure is applied 4.21 as the suspected value in the Y column.

If there is a need to use this procedure to exclude more than one value in ten, one would be cautioned to stop and to investigate the measurement process for unsuspected disturbances which may be producing the outlying values. Of course, one always has the option to run the test again.

CONCLUSIONS

The method of analysis presented in this report affords a simple, and easy to apply, criterion for objectively characterizing the quality of an intercomparison type measurement process.

The method described allows the measurement process to be objectively evaluated by computing the process W_1, W_2 and W_3 intervals.

As the quality of the process improves, the W intervals will decrease. If i.e. in subsequent tests the W intervals increase, this would indicate a decrease in the quality of the process. For example variations in procedures, variations in operator characteristics, variations in equipment functional characteristic and variations in the laboratory environmental conditions could be objectively evaluated by monitoring the magnitude of the variations in the process W intervals. See Figure 3.

One is cautioned that, to indiscriminately throw away one or two data points, introduces a subjective operator's bias into the measurement process which in-turn reduces the credibility and integrity of the process. Consequently, if there is a need to throw away one or two data points, all participating operators are advised to always use an objective data rejection method of the type described in this report.

This method of evaluation applies equally to:

(1) intercomparison tests in the same laboratory by comparing W intervals of previous tests to W interval of current tests and

(2) intercomparison tests in different laboratories by comparing the W intervals computed by one laboratory to the W intervals computed by the other laboratory as long as both laboratories have measured the same transfer standard and have agreed to compare at a specified confidence interval i.e. 95%.

References

1. Experimental Statistics; National Bureau of Standard Handbook 91; 1963 by Mary Gibbons Natrella, U.S. Govt Printing Office, Wash. D.C.

2. Precision Measurement and Calibration; Statistical Concepts and Procedures; National Bureau of Standards Handbook 300 Volume 1; 1969 by Harry H. Ku, U.S. Govt Printing Office, Wash. D.C.

TIME PHASE DIAGRAM

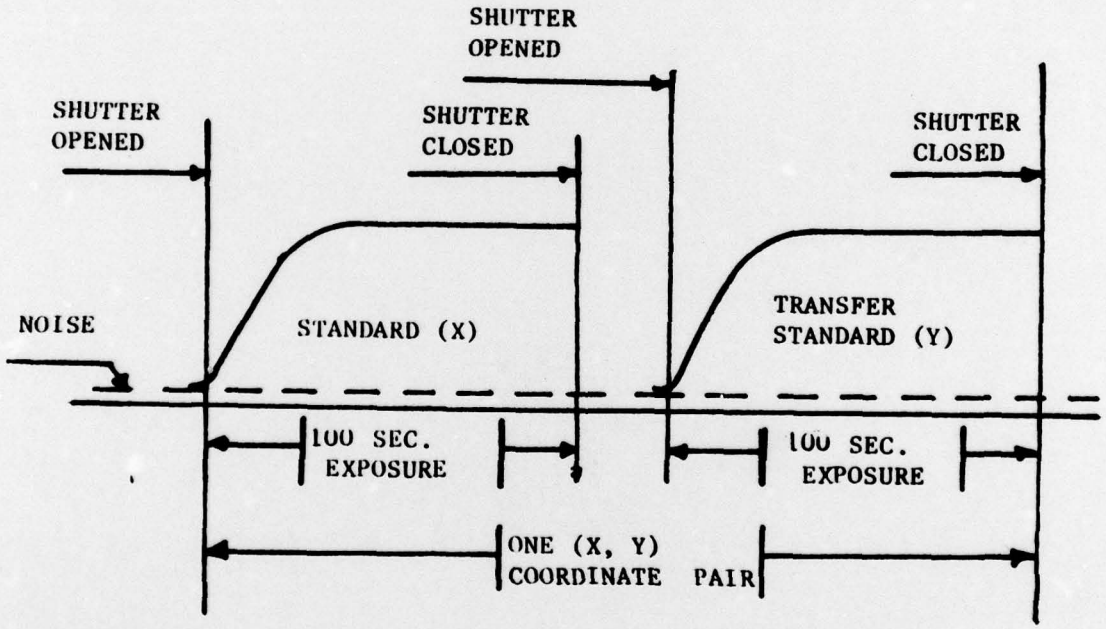


Figure 1

CURVE PLOT

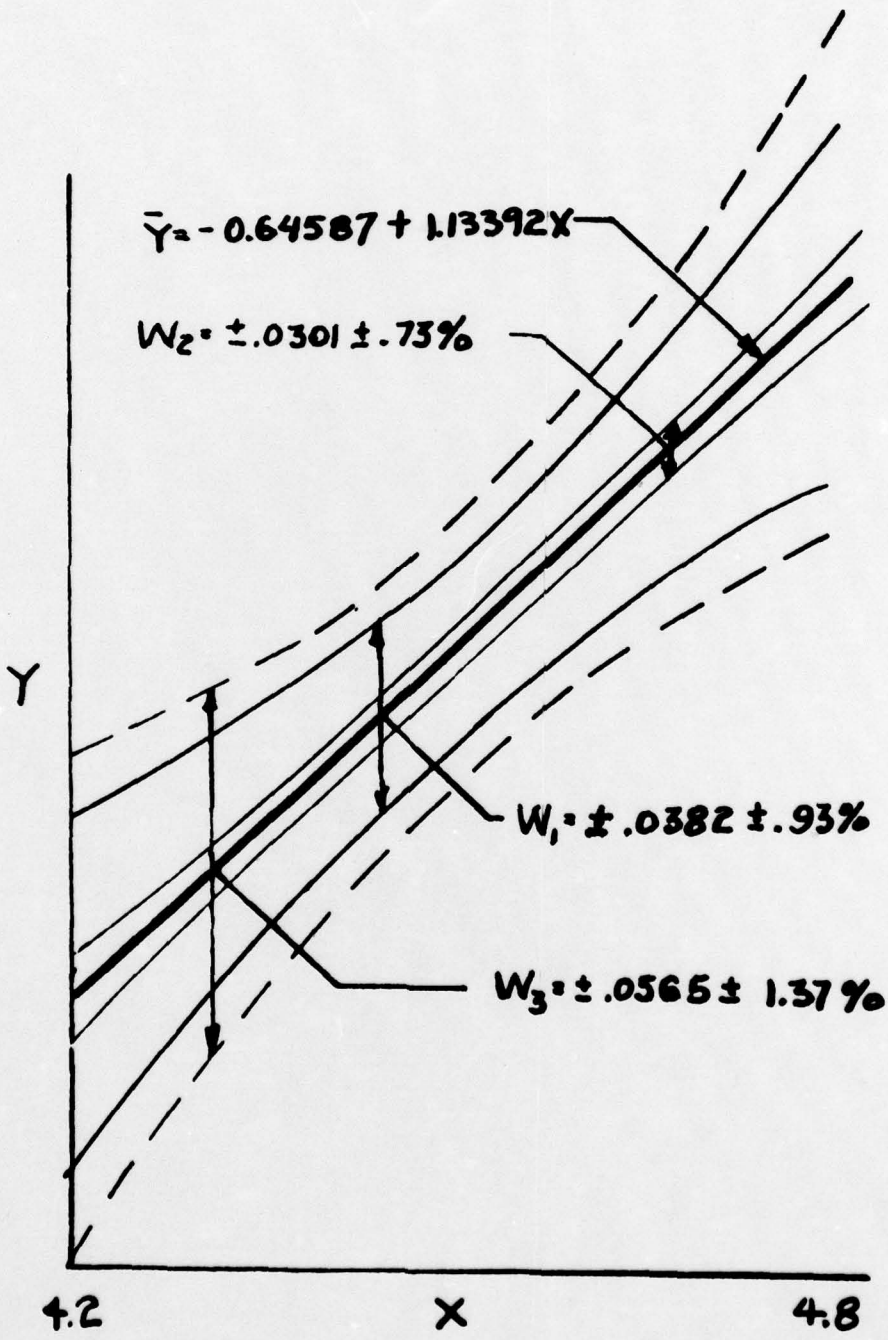


Figure 2

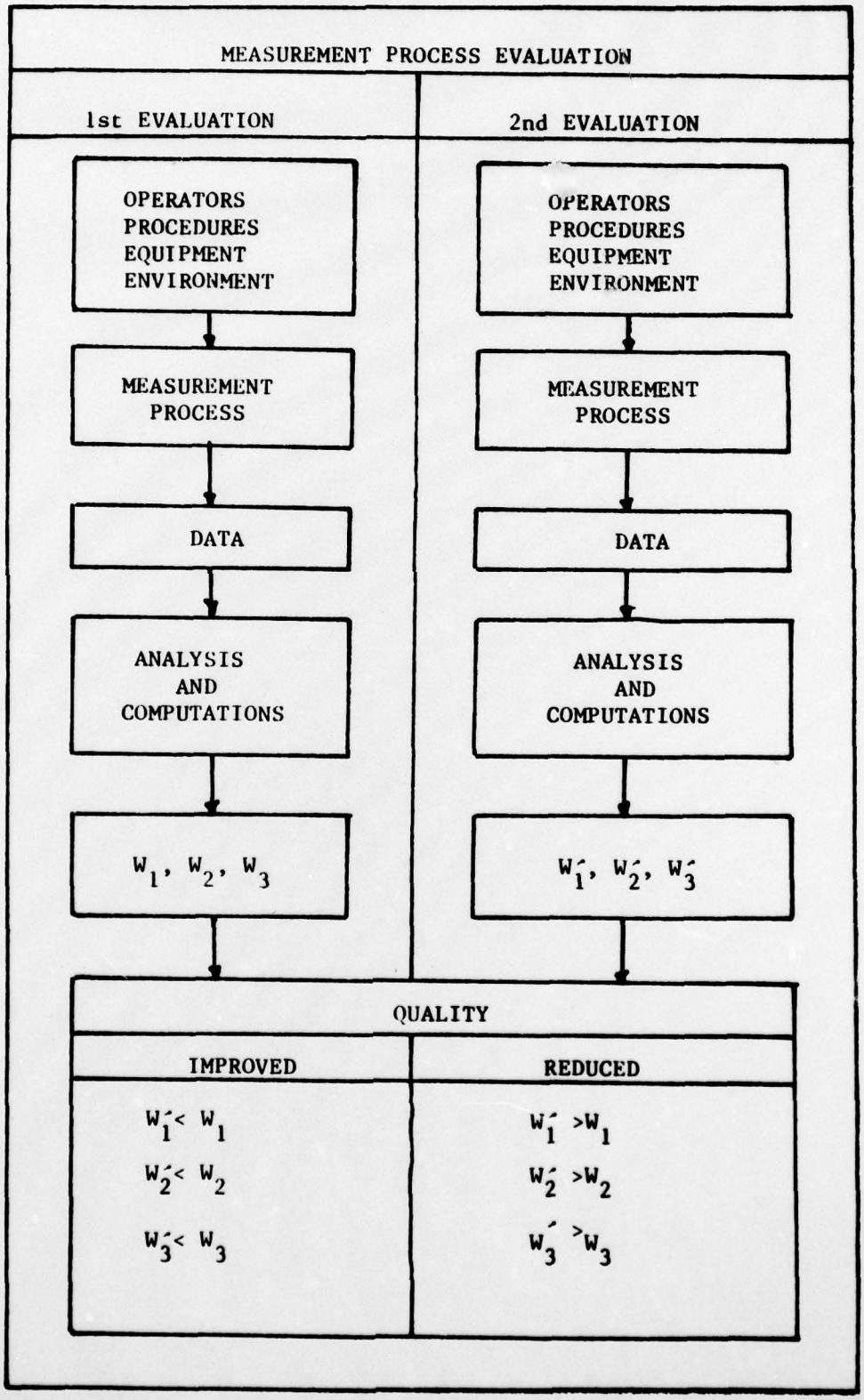

MEASUREMENT PROCESS EVALUATION
1st EVALUATION
2nd EVALUATION
OPERATORS
PROCEDURES
EQUIPMENT
ENVIRONMENT
OPERATORS
PROCEDURES
EQUIPMENT
ENVIRONMENT
MEASUREMENT PROCESS
MEASUREMENT PROCESS
DATA
DATA
ANALYSIS AND COMPUTATIONS
ANALYSIS AND COMPUTATIONS
W_1, W_2, W_3
W_1', W_2', W_3'
QUALITY
IMPROVED
REDUCED
$W_1' < W_1$
$W_2' < W_2$
$W_3' < W_3$
$W_1' > W_1$
$W_2' > W_2$
$W_3' > W_3$

Figure 3

are known to have a limited range of values of X (which is an SII Relationship).

Table 5-1 gives a brief summary characterization of SI and SII Relationships. Detailed problems and procedures with numerical examples are given for SI relationships in Paragraph 5-5.1 and for SII relationships in Paragraph 5-5.2.

BASIC WORKSHEET FOR ALL TYPES OF LINEAR RELATIONSHIPS

X denotes _____ Y denotes _____

ΣX = _____ ΣY = _____

$\bar{X}$ = _____ $\bar{Y}$ = _____

Number of points: n = _____

Step (1) ΣXY = _____

(2) $(\Sigma X)(\Sigma Y)/n$ = _____

(3) S_{xy} = Step (1) − Step (2)

(4) ΣX^2 = _____ (7) ΣY^2 = _____

(5) $(\Sigma X)^2/n$ = _____ (8) $(\Sigma Y)^2/n$ = _____

(6) S_{xx} = Step (4) − Step (5) (9) S_{yy} = Step (7) − Step (8)

(10) $b_1 = \frac{S_{xy}}{S_{xx}}$ = Step (3) ÷ Step (6) (14) $\frac{(S_{xy})^2}{S_{xx}}$ = _____

(11) $\bar{Y}$ = _____ (15) $(n-2)\,s_Y^2$ = Step (9) − Step (14)

(12) $b_1\bar{X}$ = _____ (16) s_Y^2 = Step (15) ÷ $(n-2)$

(13) $b_0 = \bar{Y} - b_1\bar{X}$ = Step (11) − Step (12) s_Y = _____

Equation of the line:

$Y = b_0 + b_1X$

_____

s_{b_1} = _____

s_{b_0} = _____

Estimated variance of the slope:

$s_{b_1}^2 = \frac{s_Y^2}{S_{xx}}$ = Step (16) ÷ Step (6)

Estimated variance of intercept:

$s_{b_0}^2 = s_Y^2\left\{\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right\}$ = _____

Note: The following are algebraically identical:

$S_{xx} = \Sigma(X - \bar{X})^2$; $S_{yy} = \Sigma(Y - \bar{Y})^2$; $S_{xy} = \Sigma(X - \bar{X})(Y - \bar{Y})$.

Ordinarily, in hand computation, it is preferable to compute as shown in the steps above. Carry all decimal places obtainable—i.e., if data are recorded to two decimal places, carry four places in Steps (1) through (9) in order to avoid losing significant figures in subtraction. "Copied from NBS Handbook 91 pg 5-10, U.S. Govt Printing Office, Washington, D.C."

APPENDIX - B

TABLE A-5 (Continued). PERCENTILES OF THE F DISTRIBUTION

$F_{.95}(n_1, n_2)$

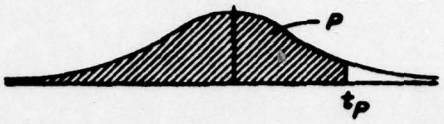
n_1 = degrees of freedom for numerator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Adapted with permission from *Biometrika Tables for Statisticians*, Vol. I (2d ed.), edited by E. S. Pearson and H. O. Hartley, Copyright 1958, Cambridge University Press.

"Copied from NBS Handbook 91 pg T-7, U.S. Govt Printing Office, Wash. D.C." for convenience of reader

TABLE A-4. PERCENTILES OF THE t DISTRIBUTION



df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576

Adapted by permission from *Introduction to Statistical Analysis* (2d ed.) by W. J. Dixon and F. J. Massey, Jr., Copyright, 1957, McGraw-Hill Book Company, Inc. Entries originally from Table III of *Statistical Tables* by R. A. Fisher and F. Yates, 1938, Oliver and Boyd, Ltd., London."
"Copied from NBS Handbook 91 pg T-5, U.S. Govt Printing Office, Wash. D.C."

APPENDIX - D

Reprinted from AMERICAN JOURNAL OF PHYSICS, Vol. 21, No. 7, 520-525, October, 1953
Printed in U. S. A.

Rejection of Outlying Observations

FRANK PROSCHAN*
National Bureau of Standards, Washington, D. C.
(Received November 24, 1952)

This paper makes available to the physicist two of the modern statistical tests for possible rejection of outlying observations. These two methods have been selected because they apply in a majority of the actually occurring situations and because they are so easy to use.

A PERENNIAL problem vexing the experimenter is that of rejection of suspected data. For one hundred years attempts at the solution of this problem have been advanced, most of them to be themselves rejected as suspect. Fortunately, modern statistical theory has proposed useful, reliable methods for objectively rejecting deviant values. However, the solution is far from complete at present.

This paper makes available to the physicist two of the modern statistical tests for possible rejection of outlying observations. These two methods have been selected because they apply in a majority of the actually occurring situations and because they are so easy to use.

THE PROBLEM

Here is a common problem facing experimenters. The typical scientist, X. Perry Menter, makes a number (say five) of repeated measurements of some unknown quantity. The smallest value (or the largest) is so far removed from the other four that he suspects that it may be in error. However, Perry has no specific knowledge that a mistake actually did occur. Let us assume that he has no previous data from which to estimate the precision of measurement. How can he decide *from the values themselves* whether the suspected value is in error or not?

The answer seems clear. He should consider the suspected value as in error when it seems too far from the other four values. But how can he judge when it is "too far from the other four values"?

A LOGICAL APPROACH

Here is a simple, logical, objective criterion. Suppose Perry could somehow make millions of sets of five observations each. Suppose, too, that he could guarantee that none of these observations had any mistakes. Call a typical set x_1, x_2, x_3, x_4, x_5, where the x's are arranged in order of size, so that $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$. Now a logical measure of the distance between the smallest value and the other four values is

$$r_{10} = \frac{x_2 - x_1}{x_5 - x_1},$$

i.e., the ratio of the interval between the suspected and adjacent value to the total range.

Now Perry records with what frequency, among his millions of sets of five values each, different values of r_{10} occur. He finds that a value of r_{10} larger than 0.780 occurs one time in one hundred. He then reasons this way:

"I have found that among sets of five observations each (*containing no mistakes*) a value of r_{10} larger than 0.780 is quite unlikely (occurs only once in one hundred). If now, in my future experiments I get a set of five observations for which r_{10} is larger than 0.780, I will conclude that my largest observation is in error."

CONFIDENCE IN THE TEST

This seems reasonable. But what confidence can Perry have in such a procedure? How often will he consider as mistaken a perfectly good observation? How often will he consider acceptable an incorrect observation?

Clearly, from the way in which he derived the test, he will classify a perfectly good smallest observation as mistaken once among one hundred sets of five each, on the average. But there is no general answer to the question of how often he will let pass a mistaken observation. This depends on how "mistaken" the mistaken observation is. If a very *large* error were made, his

* Now at Sylvania Electric Products, Inc., Hicksville, New York.

"Article copied from NBS Handbook 300 pg 349-520, U.S. Govt Printing Office, Washington, D.C."

test would tend to reject the observation almost certainly. If a very small error were made, his test would tend to reject the observation with a small probability.

Figure 1 gives some idea of the performance of r_{10} in detecting mistaken observations. It is based on a sampling experiment in which samples of five from a normal population with mean μ and standard deviation σ were contaminated with values drawn from a normal population with mean $(\mu+\lambda\sigma)$ and standard deviation σ. The ordinate shows the percent discovery of contaminators (the proportion of the time the contaminating population provides an extreme value and the test discovers this value) while the abscissa shows λ, the magnitude of the shift (error) of the contaminator in standard deviations.

We said above that once in every 100 sets of values (on the average) Perry would consider as mistaken a perfectly good observation. If he were to reject this observation and then compute the mean and standard deviation of the remaining values, these would be biased estimates. In addition, when a good observation is rejected, any further statistical tests of significance will become less reliable. This is the price that he must pay for improving the data in the cases where a mistaken observation is removed.

MATHEMATICAL DERIVATION

Of course, 0.780, the value of r_{10} that is exceeded by chance 1 percent of the time (called the 1 percent level of significance of r_{10}), is not determined by actually making millions of sets of five observations each. Rather it may be calculated mathematically[1] with even greater accuracy than if millions of sets of five observations had been used. The basic assumption is that the repeated measurements would follow the normal distribution.

LARGER SAMPLE SIZES

For sample sizes larger than seven, slight modifications in the r_{10} statistic result in a more

[1] Dixon, Ann. Math. Stat. 22, No. 1, 68–70 (1951).

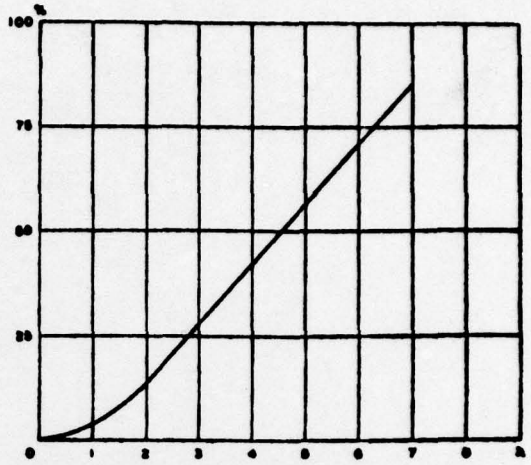


FIG. 1. Performance of r test. The ordinate shows the percent discovery of contaminators, while the abscissa shows λ, the magnitude of the shift (error) of the contaminator in standard deviations. From W. J. Dixon, Ann. Math. Stat. 21, No. 4, 493 (1950).

sensitive test. Thus for sample size $n = 8$, 9, or 10,

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$$

is superior to r_{10}. Similarly for $n = 11$, 12, or 13,

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$$

is superior. Finally for $n = 14, 15, \cdots, 30$,

$$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$$

is best.

USE OF TABLE I

Let us now define r as the appropriate statistic among r_{10}, r_{11}, r_{21}, and r_{22} according to the sample size. Table I gives critical values of r for significance levels $\alpha = 5$ percent and $\alpha = 1$ percent, for sample sizes from $n = 3$ to 30.

Thus for example for $n = 8$ and $\alpha = 5$ percent, the table gives a critical value for r (in this case r_{11}) of 0.554. This means that in 100 sets of 8 observations each, free of mistakes, five values of r_{11} will be larger than 0.554, on the average.

What if Perry suspects the acceptability of the largest observation in a set? In this case, he simply considers the observations as numbered in the reverse order and proceeds as before.

TABLE I. Testing for extreme observation (no past data).*

Statistic	Sample size n	Critical value α = 5 percent	Critical value α = 1 percent
$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$	3	0.941	0.988
	4	0.765	0.889
	5	0.642	0.780
	6	0.560	0.698
	7	0.507	0.637
$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	8	0.554	0.683
	9	0.512	0.635
	10	0.477	0.597
$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	11	0.576	0.679
	12	0.546	0.642
	13	0.521	0.615
$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	14	0.546	0.641
	15	0.525	0.616
	16	0.507	0.595
	17	0.490	0.577
	18	0.475	0.561
	19	0.462	0.547
	20	0.450	0.535
	21	0.440	0.524
	22	0.430	0.514
	23	0.421	0.505
	24	0.413	0.497
	25	0.406	0.489
	26	0.399	0.486
	27	0.393	0.475
	28	0.387	0.469
	29	0.381	0.463
	30	0.376	0.457

* By permission from W. J. Dixon and F. J. Massey, *Introduction to Statistical Analysis* (McGraw-Hill Book Company, Inc., New York, 1951), p. 319.

Why are two significance levels given? The reason is that no one significance level is appropriate to all problems. For example, consider these two cases:

(a) Additional observations are not possible.
(b) Additional observations are possible.

In case (a) for many problems it might be appropriate to compute r and test it at the 1 percent level of significance. If the particular observed value of r is larger than the tabulated value for $\alpha = 1$ percent, it might then be a good idea to exclude that observation.

In case (b), for many situations a reasonable procedure might be to test r at the 5 percent level. If the sample value of r is significant at the 5 percent level, one or more additional observations would be taken. If the observation originally suspected remained outlying, it would be tested again, using the combined set of observations. This time, however, the r test would be performed at the *1 percent* level of significance. If the outlier were significantly deviant at the 1 percent level, it would be rejected. It should be noted that among many sets tested in this way, the proportion of sets in which a perfectly good largest value will thus be rejected will be less than 1 percent. This is because the observation has a "second chance" before it is finally rejected.

SUMMARY

A set of n observations is made. No previous data are available from which to estimate the variability of a measurement. What is a rational procedure for testing whether the largest (or smallest) of the set is too deviant to be explained by the ordinary errors of measurement?

Rank the n observations in order of size from smallest to largest, if the smallest observation is suspected,

$$x_1 \leq x_2 \leq \cdots \leq x_n;$$

reverse the numbering system if the largest is suspected.

Next compute

$$r_{10} = \frac{x_2 - x_1}{x_n - x_1} \quad \text{if } n = 3 \text{ to } 7$$

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad \text{if } n = 8 \text{ to } 10$$

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1} \quad \text{if } n = 11 \text{ to } 13$$

$$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad \text{if } n = 14 \text{ to } 30.$$

Table I may be used to determine how likely it is to get as large a value of r as actually obtained, simply by chance. A procedure that might be appropriate for many problems is as follows.

(a) *No additional observations possible.* In this case, compared the computed r with the value in Table I at the 1 percent level. If the computed value of r is larger than the tabulated value, exclude the deviant observation. Otherwise, do not.

(*b*) *Additional observations possible.* In this case, compare the computed r with the value of r at the 5 percent level. If the computed value of r is larger than the value, take one or more additional observations. Otherwise accept the suspected value without taking additional observations.

If, in the enlarged set (containing all the original and the additional observations), the previously suspected value remains outlying, compute r for the enlarged set. This time compare it with the value at the 1 percent level. If the computed value exceeds the table value, exclude the outlier; otherwise do not.

EXAMPLES

1. In a preliminary experiment, Silas N. Tist makes 5 determinations of the velocity of light in vacuum by a new method, obtaining 299 792, 299 780, 299 795, 299 786, 299 820, (km/sec). Si N. Tist suspects the last value, 299 820, as being mistaken since it is so much larger than the other values. Before going on with additional experimentation, Si wishes to decide whether 299 820 is mistaken or not. What shall he do?

Since no previous data are available from which to compute the precision of measurement by this new method, the r test is appropriate. The first step is to arrange the five values in order of size: 299 780, 299 786, 299 792, 299 795, 299 820. Then

$$r = r_{10} = \frac{299\,820 - 299\,795}{299\,820 - 299\,780} = \frac{25}{40} = 0.625.$$

Since this is less than 0.780, the 0.01 point of r for $n = 5$, Si N. Tist concludes that 299 820 is not mistaken.

2. Using the Atwood machine, Norris G. Neer makes determinations of g, the acceleration of gravity, in his college course in experimental physics. N. G. Neer's values are: 986, 964, 989, 1000, 987, 909, 999 (cm/sec^2). He suspects 909 as being inconsistent with the other values. Shall he accept it, or shall he experiment further?

He computes

$$r = r_{11} = \frac{x_2 - x_1}{x_7 - x_1} = \frac{964 - 909}{1000 - 909} = \frac{55}{91} = 0.604.$$

This value lies between the 0.01 and the 0.05 points of r for $n = 7$. Hence N. G. Neer makes an additional determination and gets a new value of 971.

Since 909 remains outlying in the enlarged set of eight, he computes r for this set of eight. Now $r(r_{11})$ is 0.611. Since it is smaller than the 1 percent level of r for $n = 8$, N. G. Neer accepts 909 and uses all eight values.